

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 02/06/2006		2. REPORT TYPE Published Journal Article		3. DATES COVERED (From - To) Feb 03 – Jan 06	
4. TITLE AND SUBTITLE Fatigue Estimation Using Voice Analysis				5a. CONTRACT NUMBER F33615-03-C-6334	
				5b. GRANT NUMBER N/A	
				5c. PROGRAM ELEMENT NUMBER 65502F	
6. AUTHOR(S) Harold P. Greeley*, Joel Berg*, Eric Friets*, John Wilson,* Glen Greenough, ^Δ Joseph Picone, [√] Jeffrey Whitmore, ^Δ and Thomas Nesthus [#]				5d. PROJECT NUMBER 3005	
				5e. TASK NUMBER HP	
				5f. WORK UNIT NUMBER 08	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) *Creare, Inc. Hanover, New Hampshire ΔDartmouth-Hitchcock Medical Center Lebanon, New Hampshire √University of Mississippi Starksville, Mississippi ΔAir Force Research Laboratory Brooks City-Base, Texas #FAA Civil Aerospace Medical Institute Oklahoma City, Oklahoma				8. PERFORMING ORGANIZATION REPORT NUMBER N/A	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Materiel Command Air Force Research Laboratory 711 Human Performance Wing Human Effectiveness Division Biosciences and Performance Division Biobehavior, Bioassessment & Biosurveillance Branch Brooks City-Base, TX 78235				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/HEP	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-HE-BR-JA-2006-0002	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited , case file no. 06-038, 2/6/06. Published in Behavior Research Methods, 2007, 39(3), 610-619.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT In the present article, we present a means to remotely and transparently estimate an individual's level of fatigue by quantifying changes in his or her voice characteristics. Using Voice analysis to estimate fatigue is unique from established cognitive measures in a number of ways: (1) speaking is a natural activity requiring no initial training or learning curve, (2) voice recording is a unobtrusive operation allowing the speakers to go about their normal work activities, (3) using telecommunication infrastructure (radio, telephone, etc.) a diffuse set of remote populations can be monitored at a central location, and (4) often, previously recorded voice data are available for post hoc analysis. By quantifying changes in the mathematical coefficients that describe the human speech production process, we were able to demonstrate that for speech sounds requiring a large average air flow, a speaker's voice changes in synchrony with both direct measures of fatigue and with changes predicted by the length of time awake.					
15. SUBJECT TERMS Fatigue estimation, voice analysis, human speech production					
16. SECURITY CLASSIFICATION OF: U			17. LIMITATION OF ABSTRACT U	18. NUMBER OF PAGES 9	19a. NAME OF RESPONSIBLE PERSON Jeff Whitmore
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code)

Fatigue estimation using voice analysis

HAROLD P. GREELEY, JOEL BERG, ERIC FRIETS, AND JOHN WILSON
Creare Incorporated, Hanover, New Hampshire

GLEN GREENOUGH
Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire

JOSEPH PICONE
University of Mississippi, Starkville, Mississippi

JEFFREY WHITMORE
Air Force Research Laboratory, Brooks City-Base, Texas

AND

THOMAS NESTHUS
FAA Civil Aerospace Medical Institute, Oklahoma City, Oklahoma

In the present article, we present a means to remotely and transparently estimate an individual's level of fatigue by quantifying changes in his or her voice characteristics. Using Voice analysis to estimate fatigue is unique from established cognitive measures in a number of ways: (1) speaking is a natural activity requiring no initial training or learning curve, (2) voice recording is a unobtrusive operation allowing the speakers to go about their normal work activities, (3) using telecommunication infrastructure (radio, telephone, etc.) a diffuse set of remote populations can be monitored at a central location, and (4) often, previously recorded voice data are available for post hoc analysis. By quantifying changes in the mathematical coefficients that describe the human speech production process, we were able to demonstrate that for speech sounds requiring a large average air flow, a speaker's voice changes in synchrony with both direct measures of fatigue and with changes predicted by the length of time awake.

The unique characteristics of the military and aviation environments make war fighters, pilots, and air traffic control personnel particularly susceptible to fatigue. Environmental factors such as movement restriction, poor airflow, low light levels, background noise, and vibration are known to cause fatigue (Mohler, 1996). In addition, the introduction of advanced automation has changed the nature of the job for these individuals. "Hands-on" activities have been replaced by greater demands on the crew to perform vigilant monitoring of automated systems; a task that people find tiring if performed for long periods of time (Colquhoun, 1976). Personnel operating at unacceptable levels of cognitive performance present a danger to their mission, to themselves, and to their work team.

An analysis of NASA's Aviation Safety Reporting System (ASRS) revealed that 3.8% of air transport crew member error reports were directly associated with fatigue (Lyman & Orlady, 1981). However, when factors related to fatigue are considered, such as inattention or miscommunication, the number increases to 21.1%. Fatigue also results in an increase in what a person might consider acceptable risk in an attempt to avoid additional

effort (Barth, Holding, & Stamford, 1976; Shingledecker & Holding, 1974).

The ability to quickly and unobtrusively monitor an airman's or soldier's level of alertness prior to and during the undertaking of mission-critical activity would provide commanders with critical information regarding personnel assignments, quite possibly save lives, and increase the likelihood of mission success. Unfortunately, there are no cognitive assessment tests that have been proven to be effective in the field under conditions of high stress and severely limited time.

In this article we will describe and evaluate the application of a speech-based approach to estimating a speaker's level of fatigue. Using the voice characteristic metrics that are necessary in the implementation of most automatic speech recognition (ASR) software algorithms, we quantify the change in speakers' voice quality as they become fatigued. These changes are compared to widely accepted empirical and model-based measures of fatigue. The next section discusses the physiology of speech production as an introduction to our approach and includes previous attempts to relate voice characteristics to fatigue.

H. P. Greeley, hpg.ra@verizon.net

Human Speech Production and Its Association With Fatigue

Recognizable speech is produced by a continuous adjustment of the resonating characteristics of the vocal tract. This system consists of an excitation region (lungs, diaphragm, and vocal folds) and a filter that is adjusted by changes in the position of the pharynx, tongue, lips, jaw, and soft palate.

The production of speech sounds is a process that relies on precise interactions between the sensory and motor systems. Control of the voice articulators is done through a biofeedback process involving the sensing and monitoring of the vibration of the vocal folds through the sound and feeling that they create. With increasing fatigue or alcohol-induced impairment, this system is disrupted. Different speech-based manifestations of this disruption have been reported by a number of researchers. Previous work associating changes in voice with fatigue has generally focused on discrete characteristics of the speaker's voice. These include, pitch and word duration (Whitmore & Fisher, 1996) and the timing between articulated sounds (Vollrath, 1994). Changes in voice spectral parameters have been associated with alcohol-related impairment (Brenner & Cash, 1991) and hypoxia (Saito et al., 1980). The significant effects of circadian influences on voice characteristics have been observed in a number of studies (Roth et al., 1989; Whitmore & Fisher, 1996).

In our analytical procedure, we monitored the organized collection of sounds of the International Phonetic Alphabet (IPA), a listing of 41 phones from which all English words are comprised. In this manner, identical words need not be present in the recorded or "online" vocalizations, and more subtle changes may be detected compared to whole word analysis. Quantified speech signals using metrics that are representative of the entirety of the speaker's voice were also utilized. This process is described in the next section.

Quantification of Voice

Mathematically, the speech signal consists of a convolution of the excitation waveform with the filter description in the time domain or by a multiplication of the transfer functions of the two regions in the frequency domain. It is possible to process the recorded speech signal $[S(\omega)]$ in a manner that will separate the isolated filtering effects $[F(\omega)]$ from the excitation signal $[E(\omega)]$. In this process, the spectral characteristics of the speech signal are obtained and a logarithm of the resulting amplitude is calculated. This provides a computed measure from which excitation and filter components are separated, and can be seen in Equation 1.

$$\log[S(\omega)] = \log[E(\omega)] + \log[F(\omega)]. \quad (1)$$

The resultant log magnitude spectrum is then transformed back to the time domain using a discrete Fourier transform. This process ultimately results in the formation of a discrete (and manageable) number of coefficients (called *cepstrum coefficients*) that represents separate filter and excitation signals in the time domain. It is important to note that the entire speech production process is now characterized by only these few cepstrum coeffi-

cients. Isolation of the spectral coefficients from either the excitation or filter sections is accomplished by the removal of the irrelevant cepstrum coefficients followed by another conversion to the frequency domain.

From this discussion, the entire human speech production process may be reduced and described by a manageable number of coefficients. Thereby, instead of tracking all of the fatigue-related changes in specific vocal metrics, such as pitch or duration, we can track changes in the entire speech production system with the analysis of these coefficients. Our developmental software calculates 36 mel-frequency cepstrum coefficients (MFCCs) comprised of 12 cepstral coefficients (MFCC 1–MFCC 12) and their first and second time derivatives. From this point we will refer to these 36 components as the voice vector.

These results were achieved by using speech recognition software developed for this project by the Institute for Signal and Information Processing (ISIP) at Mississippi State University. As depicted in Figure 1, this software calculates 12 cepstral coefficients along with the first and second derivatives (the Voice Vector). This voice characterization capability was then used to analyze the experimental data described in the following section.

Figure 2 illustrates how the voice vector of a speaker had changed over a 4-day period of the sleep restriction. This example represents the voice vector generated by a single subject's utterance of the sound "t." The legend identifies the correlation coefficients for the voice vector at 12, 39, and 78 h awake with the voice vector at the onset of testing (Trial 1 at 12 h awake).

EXPERIMENTAL DATA

The data provided for this study consisted of 296 speech trials acquired from 31 normal talkers from three separate experimental protocol test conditions:

Group 1 (FAA Group)

Six nonmedicated subjects reciting 31 unrelated words every 6 h over a period of 34 h of sleep loss.

As part of a larger FAA study that involved a 34-h period of sleep deprivation (Nesthus, Scarborough, & Schroeder, 1998), subjects were asked to recite a list of 31 words during 6 test times (10:00 A.M., 4:00 P.M., 10:00 P.M., 4:00 A.M., 10:00 A.M., and 4:00 P.M.). The test times were selected to represent circadian high and low points in alertness and performance. Voice was recorded using a Digital Audio Tape (DAT) recorder (TASCAM Model No. DA-P1) and a handheld microphone. Also measured during these test times were sleep onset latencies (SOL), representing an objective method for determining sleepiness. This is described in the next section. Between test sessions, these subjects participated in low arousal activities such as reading, watching TV, and schoolwork but not allowed to sleep.

Group 2 (Air Force Group)

Nine medicated and eight placebo subjects recited eight fixed phrases every 3 h during 66 h of testing. The initial data collection occurred at 1800 h on Day 1. Given a wakeup time of 0600 h, each subject experienced ap-

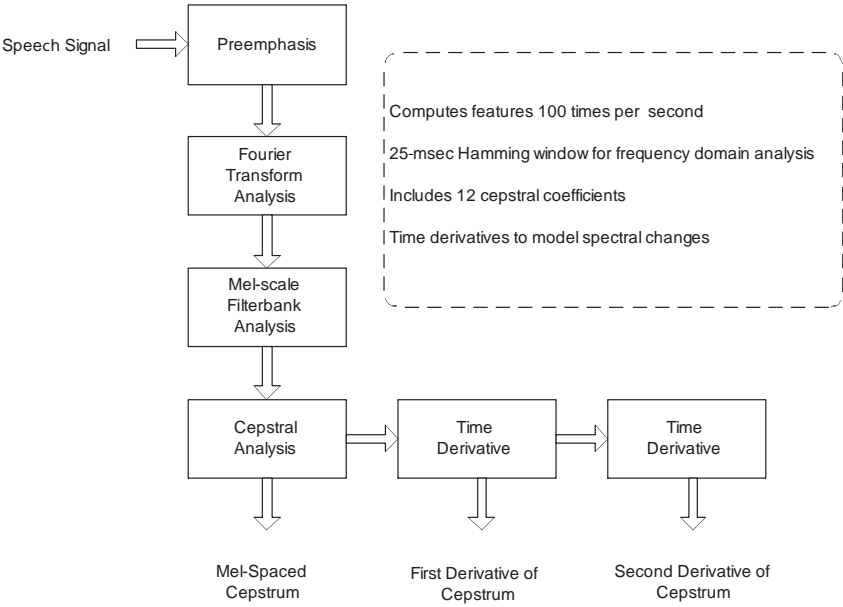


Figure 1. The information generation process used by the ISIP software. Speech signals are scaled according to human ear response characteristics using a “Mel-scale” filter bank. Cepstral analysis and two derivative processes provide 36 voiced sound coefficients for each time frame of speech data.

proximately 78 h of restricted sleep by the end of testing (1200 h on Day 3).
The objective of this effort was to evaluate the efficacy of modafinil for sustaining alertness in personnel involved in sustained field operations. All participants were males between the ages of 18 and 34 years.

Participants hiked approximately 22 miles during the first two days of the field event and then bivouacked for the remaining 24 h of the study. While traveling the route, participants performed 10 min of tests every 3 h. This test-block consisted of several simple cognitive tests (the ARES test battery) performed on a personal digital assis-

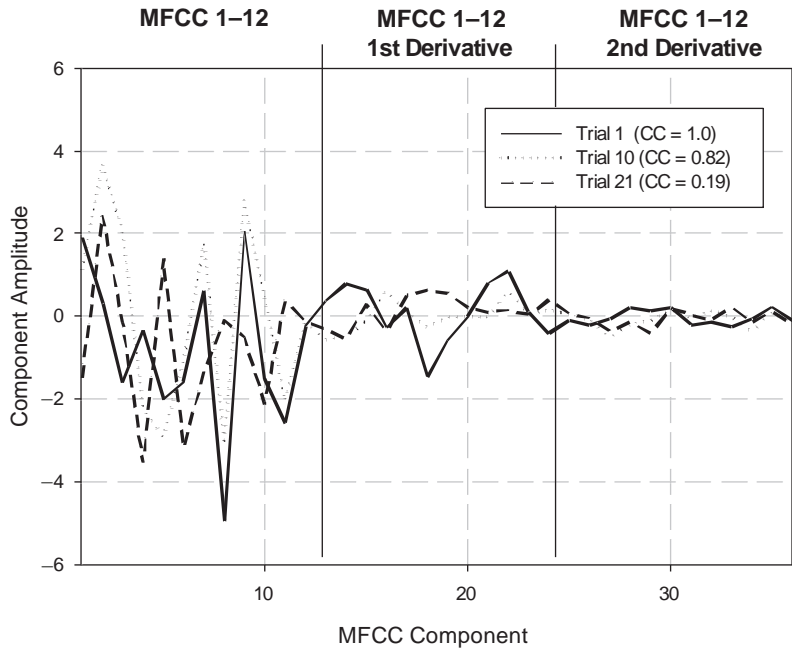


Figure 2. Changes in the voice vector during four days of sleep restriction. Here we illustrate the voice vector (36 MFCC components) generated by a single subject’s utterance of the sound “t.” As quantified in the legend, the vector during Trials 1 and 10 match better than the vectors at Trials 1 and 21.

tant (PDA) palmtop computer (Tungsten T), a subjective sleepiness check (the Stanford Sleepiness Scale), a fatigue questionnaire (the Sustained Operations Assessment Profile or SOAP), and a mood questionnaire (Profile of Mood States or POMS). Every 6 h along the route there was a checkpoint during which the normal 10-min test block was performed along with additional tests, including voice. Voice was also recorded on the PDA using that device's voice memo capability. Participants were not allowed to sleep during the first night en route and were only allowed a 2-h sleep period (0000–0200 h) during their second and third nights on the trail.

Group 3 (Creare Group)

Eight subjects recited eight fixed phrases every 2 h during a normal workday.

These volunteers were obtained from the Creare Inc. employee population. Starting with their arrival at work (between 7 and 8 A.M.), self-administered tests were conducted approximately every two h until the end of their workday (between 5 and 6 P.M.). Total time awake for each subject was calculated and recorded based upon their reported wake up times on the day of testing. During these tests, the volunteers read a series of phrases into the same model PDA used for Group 2. Between tests the subjects maintained normal work activities (office and laboratory).

The Speech Recognition Software (SRS) described in the previous section was used to process the recorded speech samples into the 36-component voice vectors, associated with the speech phones contained within each sample. These voice metrics were analyzed to determine the degree of association between the speaker's measured and estimated levels of fatigue. Figure 3 illustrates this process.

Our initial analysis revealed that the individual MFCC component most sensitive to fatigue varied from speaker to speaker. As such, comparisons of individual voice vector components would not generalize across a population of speakers. Recalling that speech recognition software can reliably recognize specific sounds spoken by a wide range of speakers by analyzing the entire voice vector, a correlation coefficient was calculated between the voice vector at Trial 1 and that obtained during the trial of interest. We call this metric the Voice Correlation or Vc. In Figure 2 we showed a speaker's voice 27 waking hours after an initial utterance (12 h awake vs. 39 h awake) had changed ($Vc = 0.82$); however, it was much closer to the "rested state" voice than to the utterance after 66 h ($Vc = 0.19$) which compares voice at 12 h awake with that at 78 h awake. Note that from this point on, the term Vc (for Voice Correlation) will be used instead of the generic correlation coefficient (CC) to emphasize that change in voice.

Changes in the resulting voice vectors were compared to physiological and behavioral measures of fatigue. This is discussed in the next section.

Performance Models and Sleepiness Measures

The hypothesis that voice changes reflect the speaker's level of sleepiness and, consequently, his or her level of performance on alertness-dependent tasks, was formulated. To test this hypothesis, three models, or sets of data, with which voice could be compared included: (1) sleep onset latency, (2) a parametric performance model, and (3) a nonparametric performance model.

Sleep onset latency (SOL). Historically, the Multiple Sleep Latency Test (MSLT) has been the primary objective test used for the measurement of sleepiness and alertness, respectively. The MSLT procedure was formalized

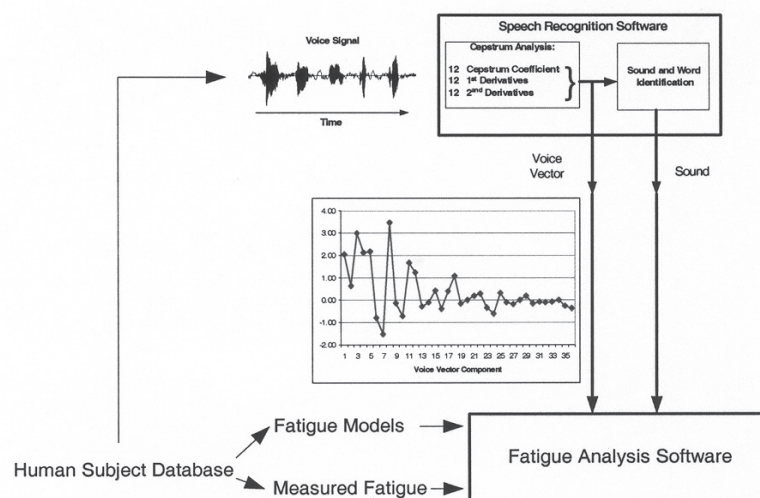


Figure 3. Performing voice quantification from the speech signal. As is common to most speech recognition software, the front end of our system recognizes individual sounds by matching the cepstral components of windowed segments of the voice signal to known sound-cepstrum templates. We refer to these cepstral components as the voice vector.

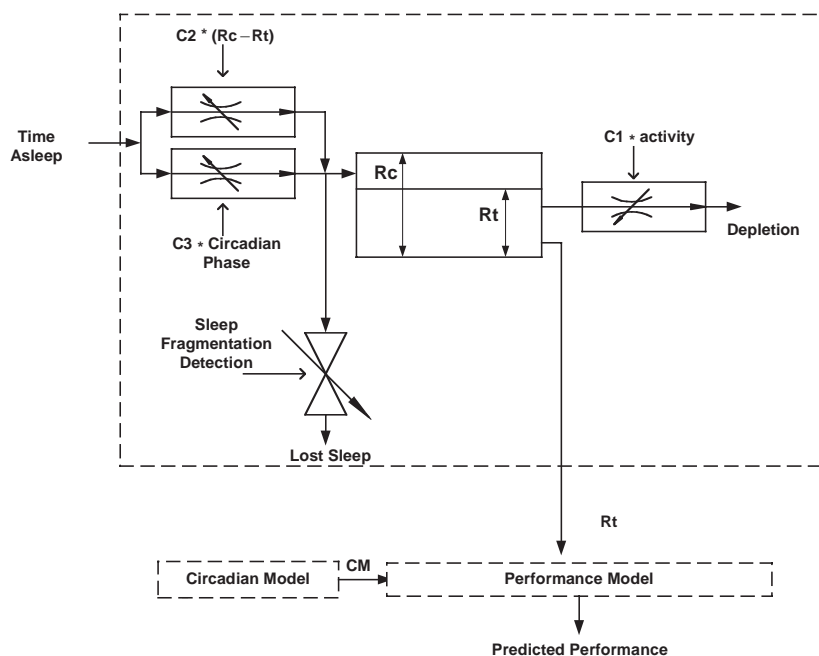


Figure 4. Block diagram of the SAFTE Model. Under fully rested conditions, a person has a finite maximal capacity to perform. This is represented by the reservoir capacity (R_c). In our depiction of this model, adjustable flow valves (shown in the upper left and center right) are used to represent flow control into and out of the reservoir. During waking activity, the reservoir is depleted. During sleep the reservoir is replenished. The constant rate terms C_1 , C_2 , and C_3 , are modeled according to each subject's test data. Circadian influences affect the rate the replenishment process and has an overall effect on performance.

in 1977 to measure sleepiness in young normal subjects involved in sleep deprivation experiments. The methodology requires subjects to be put in bed during the wake period and told to try to fall asleep. Each test is terminated after 20 min if the subject did not fall asleep. If sleep occurs, the subject is awakened after 60 sec of Stage 1 sleep. The SOL is measured from lights out to the first minute of stage 1 sleep. Significant correlations often found between the length of sleep loss and sleep latency gives face validity for using sleep onset latency as a biologically based measure of sleepiness (Arand, Bonnet, Hurwitz, Mitler, Rosa, & Sangal, 2005).

Parametric performance model. Parametric models are characterized by having a fixed structure derived by prior knowledge of the system being modeled. This prior knowledge can be taken from mathematical equations, empirical relationships, or first principles. As might be assumed, this prior knowledge requires a detailed and precise understanding of the phenomenon being modeled.

One example of a well developed parametric model of predicted performance changes with regard to sleep cycles, is the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE) model (Hursh et al., 2004). This overall parametric sleep model, illustrated schematically in Figure 4, assumes that each individual has a sleep-dependent reservoir of capacity to perform cognitive tasks. Under fully rested conditions, a person has a finite maximal capacity to perform, as represented in the figure as a reservoir value R_c . While the

individual is awake, this capacity is depleted. While asleep, the reservoir (and hence capacity) is replenished.

In a parametric model, a fixed structure, developed from an understanding of the system under study, imposes different tasks on different parts of the model. As a result, interpretation of the model's response to input parameter changes can be based on the components of the actual system. For example, the amount of replenishment that occurs during sleep is dependent upon the depth and quality of sleep, which in turn, depends upon how sleepy the individual was at sleep onset ($C_2 * [R_c - R_t]$ in Figure 4) as well as the time of sleep onset, relative to the sleeper's circadian phase. Waking during sleep produces sleep fragmentation and causes a decrease in replenishment of the reservoir. During waking hours, the amount and type of activity that the individual performs influences the amount of drain on the reservoir.

Nonparametric performance model. Nonparametric models (such as artificial neural networks) do not require a priori knowledge of the system under study and, as such, are purely data-driven. In other words, a "black box" mapping between the measured inputs and the resulting outputs is determined. While nonparametric models often perform better (from a prediction vs. measurement perspective), their lack of a physiologically-based structure makes it difficult to translate the relationship between the model's internal parameters and the system under study.

Nonparametric models can be developed based upon correlations with observed data and mathematical equa-

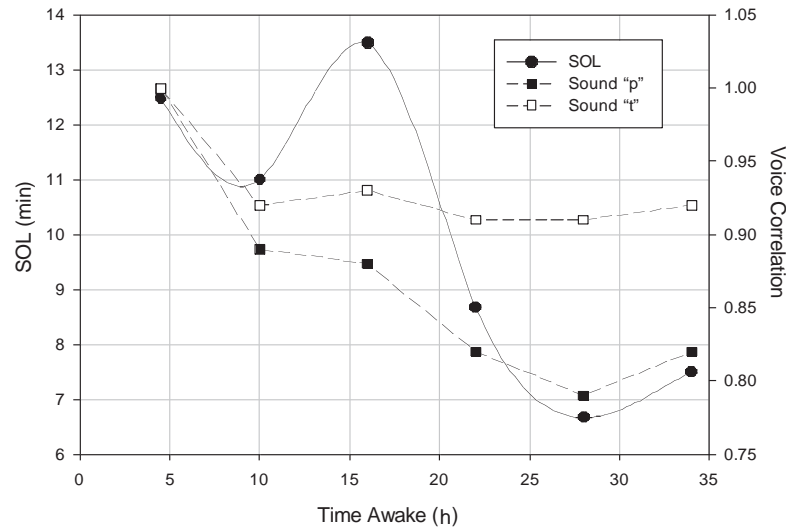


Figure 5. Change in the voice vector versus change in sleepiness. Sleep onset latency (SOL) trends downward over time for the FAA group average with circadian bumps at 16 and 28 h awake (10 P.M. and 10 A.M.). Voice correlation (Vc), is the change in the voice vector, as quantified by the correlation with the vector at Trial 1.

tions which describe this data. The resulting model's generalization capability relates to the model's ability to predict new data.

A nonparametric model was developed using an equation for passive performance suggested by Gregory et al. (2004) as shown in Equation 2. The terms on the left side of Equation 2 reflect time awake, while those on the right side reflect circadian influences. We determined the nonparametric model for each subject by adjusting the "A" and "peak" constants of Equation 2, and visually matching the model to the SOL versus time awake data.

$$\text{Passive performance (alertness)} = [(31.4 - A) * \exp(-0.527 * SD) + A] * [1 + 0.33 * \cos[6.28 * (\text{time} - \text{peak}) / 24]] \quad (2)$$

where

SD = hours awake

peak = time of day at peak performance

time = time of day

and

A is a curve fitting coefficient.

Using both SOL and Performance models as benchmarks, the present study tracked changes in the voice correlation metric (described above) as the speakers in the three test groups became fatigued. This is documented in the following section.

TEST RESULTS

Voice Change With Fatigue

Voice versus SOL. Figure 5 shows the group average change in both SOL and the Voice Correlation metric for the sounds "p" (as in pea) and "t" (as in tea) over the 34-hour sleep loss testing period in the FAA study group. It can be seen from this figure that changes in the voiced

"p" sound tracks sleepiness better than does the voiced "t" sound. It can be seen that change in articulation of the "p" sound tracks the change in sleepiness due to time awake (the abscissa of Figure 5) and is less influenced by circadian effects than the sleep onset latency. Using time awake as the independent variable, the correlation coefficients (R) between SOL and time awake is -0.825 , between $Vc(p)$ and time awake is -0.89 , and between $Vc(t)$ and time awake is -0.67 . From these numbers we estimate (using the value R^2) that time awake accounts for 68%, 79%, and 45% of the variation of SOL, $Vc(p)$, and $Vc(t)$, respectively. It can be supposed that circadian influences contribute a significant amount to the remaining variation of SOL, but less so to voice change.

Voice Versus Performance Models

Figure 6 compares the FAA study-group average SOL against both parametric (SAFTE) and nonparametric models. For each subject, a circadian model (twin circadian peaks) was determined by optimizing model output with his or her temperature (measured at the ear). Combined with sleep onset latency versus time awake data, a SAFTE model was developed.

As shown in Figure 6, over the course of 34 h of wakefulness, the nonparametric model, which is data-driven, matches the circadian pattern of the SOL much more closely than the SAFTE model.

Figure 7 compares the FAA study group's average voice changes for the sound "p" with the two models. Unlike the SOL example (Figure 6), the parametric SAFTE model follows the voice-change data more closely than the nonparametric model.

Figure 8 compares the Air Force placebo group's voiced "p" sound data with the two models. Both models appear to follow a trend in a similar manner to the voice data, though the data-driven model (nonparametric) shows a

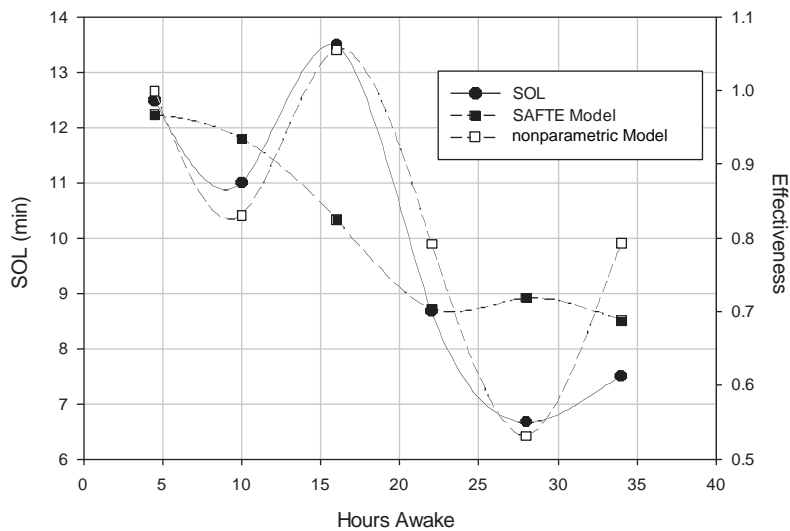


Figure 6. Performance models versus average FAA group sleepiness measurements. While all three curves trend downward with hours awake, the nonparametric model (Equation 2) shows a tighter fit to SOL than the parametric SAFTE model ($R = 0.91$ vs. 0.78).

somewhat closer match. In this instance, the correlation between the SAFTE model and $Vc(p)$ is 0.53, while that between the nonparametric model and $Vc(p)$ is 0.81.

Differences Between the Groups

The right panel of Figure 9 illustrates the voice change metrics of the voiced “p” sound for our three subject groups. The FAA group averages (filled circles) showed less change over time than did the Air Force group (filled squares). This difference might be explained by the relative level of activity performed by each of the two groups. As il-

lustrated in the SAFTE model schematic diagram (Figure 4), the rate of loss in the performance reservoir (R_t) is proportional to the level of activity while awake. The left panel of Figure 9 illustrates how changing this parameter affects the rate of performance decline for a simulated data set. The ability to perform this type of analysis is an advantage of parametric modeling over nonparametric modeling. As discussed earlier, the FAA group performed significantly less rigorous activity between testing periods than the Air Force group (daily office-type routine vs. hiking).

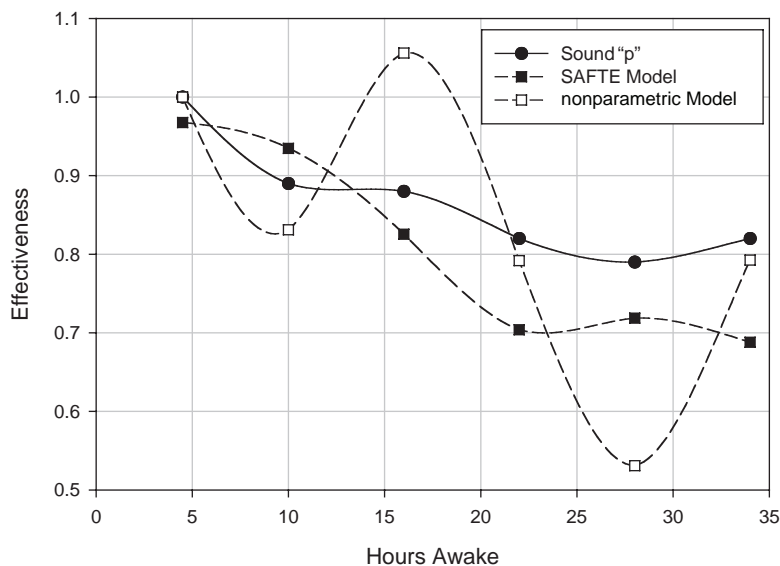


Figure 7. Change in the voice vector versus the performance models for the FAA group average. Due to the limited effect of circadian rhythms on the voice data, the SAFTE model provides a better match of the voice change pattern than the nonparametric model of Equation 2 ($R = 0.90$ and 0.73 , respectively).

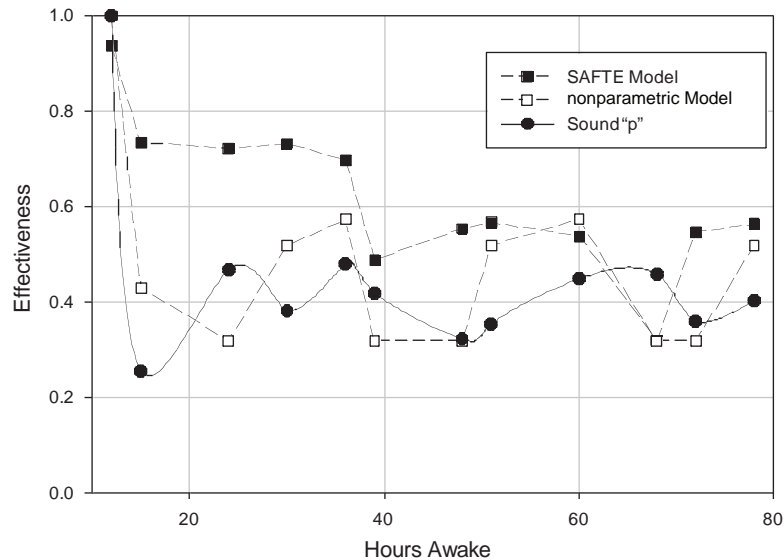


Figure 8. Change in the sound “p” voice vector versus the performance models for the Air Force placebo group average. The large change in effectiveness between the rested state (Trial 1 at 12 h awake) and Trial 21 (at 78 h awake) is predicted by all three approaches. The data-driven nonparametric model (Equation 2) provides a better match to the voice change data than the SAFTE model ($R = 0.81$ vs. 0.53).

In order to test the hypothesis that this difference is responsible for the observed differential voice decrements over time, comparisons were made between these data and that obtained from the Creare test group. This group experienced a slightly higher level of activity compared with the FAA group. As illustrated by the filled triangle symbols of Figure 9 (right panel), the rate of decline in voice data for the Creare group also shows only a slightly

greater decline in performance over time compared with the FAA group.

Differences Between the Sounds

Results show that voice sensitivity to fatigue depends, in part, upon the sound being uttered. A possible explanation for this can be based on the amount of airflow associated with each sound.

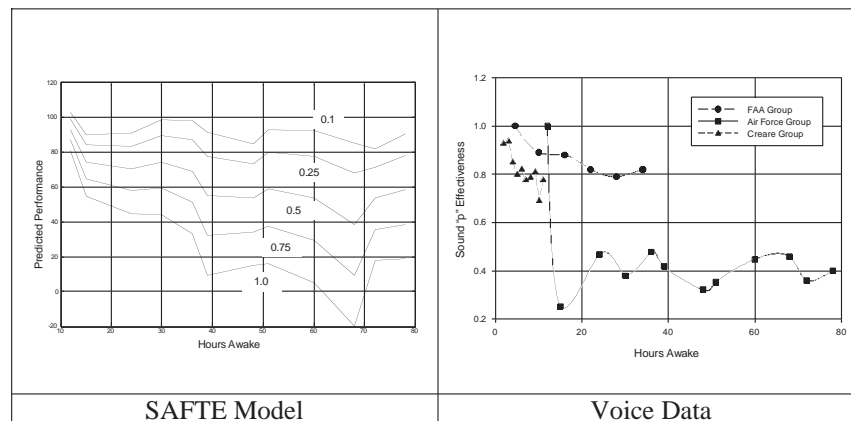


Figure 9. Rate of voice change for all three test groups. As shown in the right panel, the FAA group (closed circles) shows a much slower rate of voice change than does the Air Force placebo group (closed squares). The average change in the voice correlation for the Creare group (closed triangles) matches the FAA group much more closely than it does the Air Force placebo group. The Creare and FAA groups had a similar activity pattern over the testing period than did the Air Force group. For all groups, the voice metric is based upon the sound “p.” Change in the SAFTE model performance estimation with subject activity during periods awake is shown on the left panel. As the activity drain of the SAFTE model (C1* activity in Figure 4) increases from 0.1 to 1.0, the overall rate of performance decline increases while circadian and time awake related patterns remain the same.

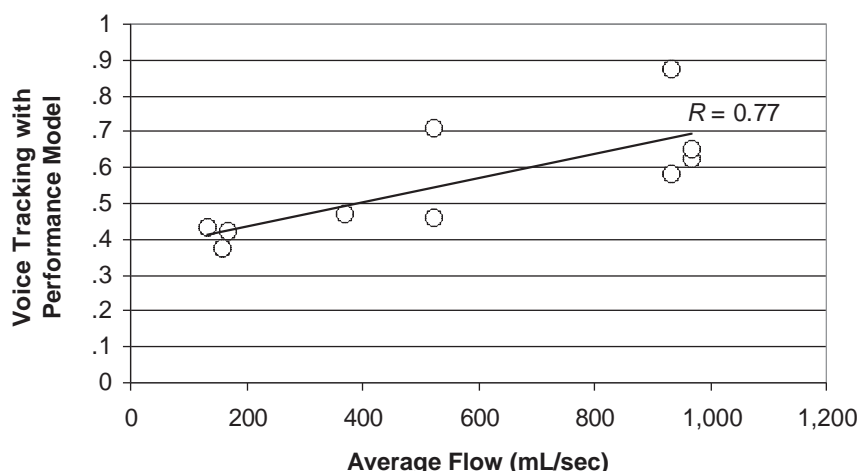


Figure 10. The association between average flow and estimation score of the Air Force and FAA voice data. Here voice correlation with the nonparametric model is plotted versus airflow. It can be seen that there is a significant correlation ($P = .01$) between the estimation capability of a sound and the average airflow required to utter that sound.

Airflow in the respiratory tract is a function of driving pressure and resistance. Driving pressure comes from the lungs and resistance is produced along the respiratory tract. Common locations for modulated resistance include the larynx and oral cavity. The relationship of airflow to driving pressure and resistance can be represented as follows (Equation 3):

$$U(\text{flow}) = P(\text{driving pressure})/Z(\text{airway resistance}). \quad (3)$$

Table 1 lists average airflow required to articulate the sounds analyzed in this study.

When one compares average airflow for specific consonants, with changes in the vocal characteristics of those consonants with fatigue, an association begins to emerge. Generally, vocal changes when verbalizing consonant sounds that require a high average airflow, were found to be more sensitive to fatigue. For example, Figure 10 illustrates the association between voice change, for both the Air Force and the FAA test groups, and the nonparametric performance model verses the average air flow for the monitored sounds. We see a significant ($P = .01$) relationship between voiced sound performance estimation ability and the average flow required to utter that sound.

DISCUSSION

Periodic speech recordings were made during three separate test protocol conditions. Generally, during (1) 2 days of testing in a relaxed setting with no sleep, (2) a 4-day hike with restricted sleep, and (3) a day of testing in a work environment. With the aid of specially designed speech recognition and voice component analysis software, these data were analyzed to isolate individual sounds (speech phones) that are most sensitive to fatigue, and to quantify the characteristics of these sounds. For particular sounds, changes in these characteristics were used to quantify and estimate the speaker's level of fatigue.

The testing conditions reported here indicate that, for speech sounds requiring a large average air flow, a speaker's voice changes in synchrony with both direct measures of fatigue and with changes predicted by the length of time awake. Comparison of voice change with models based upon time awake, such as the SAFTE model, has limitations due to the observation that time awake does not accurately quantify the speaker's level of activity over that time.

While many physiological systems experience a circadian influence, they do not do so with the same sensitivity. This appears to be the case with sleepiness (as measured by the sleep onset latency) and voice change.

Even with these differences, we believe that the association between voice change, time awake, and performance can be used as the basis for an operational setting in which remotely monitored voices can be analyzed to estimate the speaker's level of fatigue. The results presented here are a first step in this development process.

Fatigue has previously been shown to affect voice at a number of levels. This includes anatomical timing between articulation of sounds (Vollrath, 1994), time between sounds within a word (Krüger & Vollrath, 1996), and total word duration (Whitmore & Fisher, 1996). As is the case with many biological systems, circadian effects and biological individually combine to make sensitivity to fatigue individual-specific (Roth et al., 1989).

Table 1
Average Airflow Necessary to Generate the
Speech Sounds Used in Our Analysis

Sound	Average Airflow
/t/	968
/p/	933
/d/	525
/g/	372
/l/	133
/m/	168
/z/	159

Our approach to voice analysis is unique in that it does not focus on a single, discrete, voice parameter but, instead, examines changes in a mathematical representation of the entire voice (the Cepstral components). This representation is highly individual-specific; in fact, this is, in part, how voice recognizers work.

As a next step, the sensitivity and specificity of this voice-based approach should be determined by way of testing similar to that reported here with an analysis of individual (as opposed to group average) voice data from a significantly larger population of subjects.

The resulting fatigue estimation system should serve as a decision aid tool. However, due to variable sensitivity and specificity typical of most biomedical measurements, the final determination of fatigue-related effects should still be the task of a human evaluator.

AUTHOR NOTE

The material presented in this article is based on work supported by the United States Air Force under Contract F33615-03-C-6334. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the author(s) and do not necessarily reflect the views of the United States Air Force. Correspondence concerning this article should be addressed to H. P. Greeley, Creare Incorporated, Box 71, Hanover, NH 03755 (e-mail: hpg.ra@verizon.net).

REFERENCES

- ARAND, D., BONNET, M., HURWITZ, T., MITLER, M., ROSA, R., & SANGAL, R. B. (2005). The clinical use of the MSLT and MWT. *Sleep*, **28**, 123-144.
- BARTH, J., HOLDING, D. H., & STAMFORD, B. A. (1976). Risk versus effort in the assessment of motor fatigue. *Journal of Motor Behavior*, **8**, 189-194.
- BRENNER, M., & CASH, J. R. (1991). Speech analysis as an index of alcohol intoxication—The Exxon Valdez accident. *Aviation, Space, & Environmental Medicine*, **62**, 893-898.
- COLQUHOUN, P. (1976). Psychological and psychophysiological aspects of work and fatigue. *Activitas Nervosa Superior*, **18**, 257-263.
- GREGORY, J. M., XIE, X., & MENGEL, S. A. (2004, March). SLEEP (sleep loss effects on everyday performance) model. *Aviation, Space, & Environmental Medicine*, **75**, A125-A133.
- HURSH, S. R., REDMOND, D. P., JOHNSON, M. L., THORNE, D. R., BELENKY, G., BALKIN, T. J., ET AL. (2004, March). Fatigue models for applied research in warfighting. *Aviation, Space, & Environmental Medicine*, **75**, A44-A53.
- KRÜGER, H. P., & VOLLRATH, M. (1996). Temporal analysis of speech patterns in the real world using LOGOPORT. In J. Fahrenberg & M. Myrtek (Eds.), *Ambulatory assessment* (pp. S101-S113). Seattle: Hogrefe & Huber.
- LYMAN, E. G., & ORLADY, H. W. (1981, March). *Fatigue and associated performance decrements in air transport operations* (NASA Contractor Report 166167). Moffett Field, CA: NASA Ames Research Center.
- MOHLER, S. R. (1966). Fatigue in aviation activities. *Aerospace Medicine*, **37**, 722-732.
- NETHUS, T. E., SCARBOROUGH, A. L., & SCHROEDER, D. J. (1998). Changes in the performance of a synthetic work task as a function of age, gender, and sleep deprivation. In *Proceedings of the Human Factors and Ergonomics Society, 42nd Annual Meeting* (pp. 181-185).
- ROTH, T., ROEHRS, T. A., CARSKADON, M. A., & DEMENT, W. C. (1989). Daytime sleepiness and alertness. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and practice of sleep medicine* (2nd ed., pp. 14-23). Philadelphia: Saunders.
- SAITO, I., FUJIWARA, O., UTSUKI, N., MIZUMOTO, C., & ARIMORI, T. (1980). Hypoxia-induced fatal aircraft accident revealed by voice analysis. *Aviation, Space, & Environmental Medicine*, **51**, 402-406.
- SHINGLEDECKER, C. A., & HOLDING, D. H. (1974). Risk and effort measures of fatigue. *Journal of Motor Behavior*, **6**, 17-25.
- VOLLRATH, M. (1994). Automatic measurement of aspects of speech reflecting motor coordination. *Behavior Research Methods, Instruments, & Computers*, **26**, 35-40.
- WHITMORE, J., & FISHER, S. (1996). Speech during sustained operations. *Speech Communication*, **20**, 55-70.

(Manuscript received February 15, 2006;
revision accepted for publication August 3, 2006.)